DOCUMENT RESUME

ED 410 303 TM 027 577

AUTHOR	Vanneman, Alan
TITLE	New Software Makes NAEP Data User Friendly.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-97-045
PUB DATE	Apr 97
NOTE	7p.
PUB TYPE	Collected Works - Serials (022) -- Reports - Descriptive (141)
JOURNAL CIT	Focus on NAEP; v2 n3 Apr 1997
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; *Computer Software; Computer Software Evaluation; Computer Uses in Education; *Data Analysis; Data Collection; *Databases; Educational Assessment; *Educational Research; *Elementary Secondary Education; Federal Government; *Information Dissemination; Longitudinal Studies; National Surveys; State Programs; User Needs (Information)
IDENTIFIERS	*National Assessment of Educational Progress; Statistical Package for the Social Sciences

ABSTRACT

The National Assessment of Educational Progress (NAEP) has been collecting data on student achievement since 1969. It currently maintains three different assessments: long-term trends, cross-sectional national, and cross-sectional state-by-state data. Although the data are available to researchers outside the Federal Government, limited use has been made of them, due in part to the quantity and complexity of the data. The National Center for Education Statistics (NCES), which administers NAEP, has developed a number of software products to increase the accessibility and usability of NAEP data. This report gives an overview of the contents of the NAEP databases, the problems researchers face in working with them, and the software tools that have been developed to help overcome these problems. All of the NAEP databases are appropriate for assessing the proficiency of populations rather than individual students. To analyze NAEP data accurately, both sampling and the psychometric design must be taken into consideration. Researchers with the proper background can use the NCES software to work with NAEP data rapidly and efficiently. The following products are described: (1) the 1992 and 1994 "NAEP Almanac Viewers" on CD-ROM; (2) "NAEP Data on Disk Assessment Series" (CD-ROM); (3) "The NAEP Data Extraction Program," to be used with "Data on Disk"; and (4) a Statistical Package for the Social Sciences (SPSS) module to be used with SPSS for Windows. How to order these products, and how to obtain the restricted use licence needed for the "NAEP Data on Disk Assessments" is outlined. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

NATIONAL CENTER FOR EDUCATION STATISTICS

THE NATION'S REPORT CARD

Focus on NAEP

Vol. 2 No. 3 April 1997

ED 410 303

New Software Makes NAEP Data User-Friendly

The National Assessment of Educational Progress (NAEP) has been collecting data on student achievement since 1969. It currently maintains three different assessments—long-term trends, cross-sectional national, and cross-sectional state-by-state. The three assessments offer extensive data on student performance at the 4th, 8th, and 12th grades in a variety of subjects, as well as data on student and teacher background and school and classroom educational practices. Although these data are available to researchers outside the federal government, use by secondary researchers has been limited, due in part to the quantity and complexity of the NAEP data. The National Center for Education Statistics (NCES), which administers NAEP, has developed a number of software products to increase the accessibility and usability of NAEP data. In addition, NCES provides funding for secondary research projects using NAEP data. This *Focus on NAEP* will give an overview of the content of the NAEP data bases, the problems that researchers face in working with the NAEP data, and the software tools that have been developed to help overcome those problems.

The following NAEP software tools are currently available:

- *The 1992 and 1994 NAEP Almanac Viewers on CD-ROM*, which offer easy access to publicly available NAEP data for those years.
- *NAEP Data on Disk Assessment Series*, CD-ROMs that offer complete NAEP assessment data. These disks are subject to confidentiality restrictions.
- *The NAEP Data Extraction Program*, to be used with *NAEP Data on Disk*, allows researchers to extract and manipulate the particular variables of interest to them.
- *An SPSS module,* to be used with *SPSS® for Windows™,* that automatically carries out the complex analyses necessary for accurate analysis of NAEP data.

The NAEP Databases: An Overview

The National Research Council recently described NAEP data as "an unparalleled source of information about the academic proficiency of U.S. students, providing among the best available trend data on the academic achievement of elementary, middle, and secondary students in core subject areas. In addition, NAEP has distinguished itself in setting an innovative and rigorous agenda for conventional and performance-based testing." The NAEP data constitute a rich resource for research that has remained largely untapped until now, due in part to the complex-

TM027577

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ity of the technical programming needed to analyze it.

The *NAEP Long-Term Trend Assessments* cover trends in student proficiency in reading, writing, science and mathematics at ages 9, 13, and 17. (The writing long-term assessment samples by grade—4th, 8th, and 11th—rather than age.) To ensure comparability of data, these assessments rely exclusively on items that have been used in past assessments and they are administered using the same timings and instructions as past assessments. The reading, science, and math long-term assessments offer data covering a time span of more than 20 years, while the writing long-term trend assessment began in 1984. The long-term trend assessment data allow breakdowns in trend data by quartiles, race-ethnicity, gender, region, type of community, type of school, and parental education. The long-term trend assessments also collect data on trends in course-taking and trends in school and home contexts for learning.

The *NAEP Cross-Sectional Assessments*, which began in the 1980s, were developed to respond to changes in curricular emphases and objectives and to include new material that educators and other experts believe should be in the assessments, while still maintaining a connection with previous assessments. In addition to the core subjects of reading, writing, science, and mathematics, NAEP has done cross-sectional assessments in history, geography, civics, and the arts. The cross-sectional assessments offer breakdowns by the same categories as the long-term assessments, with additional data on teaching practices.

The *NAEP State Assessments* permit state-by-state comparisons. While participation is voluntary, about 40 states and territories have participated in each of the state assessments, which began in 1990. The state assessments use the same questions as the cross-sectional assessments, but different sampling designs. The state assessments use separate samples for each state, while the national assessment draws a single sample that represents the entire nation. Thus, the national cross-sectional and state assessments form separate databases.

The cross-sectional and state assessments include performance standards or achievement levels, reported against the NAEP scale. They indicate what students should be able to do. The achievement levels, which have been used in 1992, 1994, and 1996, are still considered developmental and subject to further review.

The *High School Transcript Study* is a NAEP-related database. The High School Transcript databases are available for 1982 (High School and Beyond Survey), 1987, 1990, and 1994 and allow analysis of coursetaking patterns and student achievement. The 1994 study, like the 1987 and 1990 studies, drew on the same population of 12th-graders that was sampled by the NAEP assessments for that year. In most cases, transcript data can be linked with NAEP assessment data for the same students.

All of the NAEP databases are appropriate for assessing the proficiency of *populations*, not individual students. NAEP is prohibited by law from releasing scores identifiable by student or by school. NAEP scores would in any event be inappropriate for evaluating either individual student or school performance, due to the assessment design.

Complexity of NAEP Data Analysis

Secondary researchers can obtain access to all the NAEP databases. NAEP collects data via a multi-stage, clustered sampling design involving unequal selection probabilities. Since most popular statistical analysis packages assume simple random sampling, special statistical programs are needed to analyze the NAEP data accurately. The NAEP test instrument is also built on a complex model. This model produces

scores, called plausible values, that estimate proficiency and the measurement error associated with each examinee's score. To analyze the NAEP data accurately, both the sampling and the psychometric design must be taken into account. Researchers with the proper background can use the software tools developed by NCES to work rapidly and efficiently with NAEP data.

The NAEP Software

NAEP has developed a variety of software packages to make NAEP data more accessible. Some, like the Almanac Viewers, require no statistical expertise on the part of the user, while others, like the SPSS Module, support sophisticated research projects. The various packages are described below.

The *NAEP Almanac Viewer*, part of *the NAEP Data on Disk* series, offers a DOS-based, menu-driven search system for examining NAEP cross-sectional almanac data. Almanac viewers are currently available for 1992 and 1994 assessments. Release of 1996 data almanac viewers are scheduled for late 1997. Each viewer is available on a single CD-ROM, which also contains the almanac data. The Almanac Viewers offer a crosstabulation of every student, teacher, and school background variable in NAEP by about 10 demographic variables, such as student gender, region, race-ethnicity, parental education, type of school, and type of community. NAEP average proficiency scale scores are available for each cell. For example, researchers can locate the reading scale scores for 4th-grade Hispanics living in big cities, suburbs, medium cities, and small towns. They can locate the average reading proficiency of 4th-grade Hispanics who use the school library every day, once a week, once a month, or once a year. Comparisons between Almanac data must take into account the standard errors (provided in each instance) to achieve statistical significance. Almanac data are readily comprehensible without further analysis, although they can be used for a variety of research purposes. They are *not* subject to any restrictions on use. Tables can be copied into popular word-processing programs for use in reports and other documents. Specific contents of the two viewers now available are as follows:

The *1992 NAEP Almanac Viewer* contains data for the 1992 cross-sectional assessment and the 1992 Trial State Assessment. In 1992, NAEP assessed the reading, mathematics, and writing knowledge and skills of nationally representative samples of students in grades 4, 8, and 12. NAEP also assessed representative samples of public school students from 44 states and territories in reading (grade 4) and mathematics (grades 4 and 8). The viewer includes comparable 1990 data, for mathematics only. The 1992 viewer can be used with Windows, employing the DOS window. However, an operator using a computer equipped with Windows 95 must exit this program entirely and go into the DOS mode to use the 1992 viewer.

The *1994 NAEP Almanac Viewer* contains data for the 1994 cross-sectional assessment and the 1994 State Assessment. In 1994, NAEP assessed the reading, U.S. history, and geography knowledge and skills of nationally representative samples of students in grades 4, 8, and 12. NAEP also assessed representative samples of public school students from 44 states and territories in reading (grade 4). The viewer includes comparable 1992 data, for reading only. The 1994 viewer can be used with both Windows and Windows 95 operating systems, using the DOS window. These data are also available via the Internet at the NAEP homepage (http://www.ed.gov/NCES/NAEP/) in the form of downloadable PDF files. These files can be printed out once they have been downloaded, but tend to lose formatting when copied into word-processing files.

The *NAEP Data on Disk Assessment Series* provides any secondary user, whether researcher

or policymaker, with *all* available data collected, derived, and analyzed during each assessment cycle on a single CD-ROM. These disks contain "microdata," that is, data regarding individual students and schools, which *are subject to confidentiality restrictions*. Only organizations that have obtained a license and have sworn not to disclose individual or school-level results are allowed access to these CD-ROMs. (For information on how to obtain a license, see below.) Currently, six disks are available: the 1990 National Assessment (both cross-sectional and long-term trend); the 1990 Trial State Mathematics Assessment; the 1992 National Assessment (both cross-sectional and long-term trend); the 1992 Trial State Mathematics Assessment; the 1992 Trial State Reading Assessment; and the 1994 State Reading Assessment.

The *NAEP Data Extraction Program (NAEPEX)* assists secondary users in the selection and manipulation of the many samples found in the secondary-use data files (any of the six *NAEP Data on Disk Assessment Series* CD-ROMs described above). A typical data set consists of hundreds of variables. The Data Extraction Program allows users to select and extract the variables they wish to examine. This software will also create SAS and SPSS control statements for use in the creation of system files. The current NAEPEX program is DOS-based. A Windows version is being developed.

An *SPSS module* that links into the SPSS® for Windows™ program simplifies the statistical analysis of the NAEP data. These programs are linked to SPSS versions 6.0.1 and 6.1 for Windows or any of the later versions of SPSS. This module is used to produce user-defined cross-tabular analyses and also performs regression analyses. The programs automatically estimate means and regression coefficients using all five plausible values, and take all the steps necessary to appropriately estimate the standard errors for these statistics without requiring special input from the user. A similar SAS module that links into the SAS® for Windows™ software is under development.

Other NAEP-Specific Software

Apart from these NCES-developed programs, Bryk and Raudenbush's HLM software for hierarchical linear modeling contains a subroutine that is especially adapted to working with NAEP data. HLM Version 4 is available from Scientific Software International (312–684–4920). Several NCES-funded NAEP research projects have used HLM/NAEP software successfully (see "For Further Information" below for examples).

Confidentiality Restrictions

To obtain any of the NCES-developed software described in this paper, contact Bob Clemons at 202–219–1690, or e-mail bob_clemons@ed.gov. A *restricted-use data license* is required for the use of NAEP Data on Disk Assessment Series products. Only organizations, not individuals, can obtain a restricted-use data license. To obtain a license, follow these procedures:

1. Obtain a copy of the NCES Field Restricted Use Data Procedures Manual from Cynthia Barton at 202–219–2199.
2. Prepare an abstract of your research design.
3. Determine which database(s) you wish to analyze.
4. Determine who should be authorized to use each database.
5. Design a computer security plan.

After completing these requirements, send a formal letter of request for the data to:
Data Security Assistant
Department of Education/NCES/SSMD
555 New Jersey Avenue NW
Room 408
Washington, DC 20208–5654

Use the NCES Field Restricted Use Data Procedures Manual to prepare your formal letter of

request. The letter must be printed on official organization letterhead and must contain the following information:

- the title of the database(s) you want to access;
- a description of the statistical research project necessitating access to the database(s);
- the name and title of the Senior Official;
- the name and title of the Principal Project Officer(s);
- the names and titles of the professional-technical and support staff;
- the estimated loan period (not to exceed five years); and
- the desired computer media format.

For Further Information

The following publications provide information about secondary analysis of NAEP data. To obtain NCES publications listed below, call Bob Clemons at 202–219–1690, or e-mail bob_clemons@ed.gov. For more information about the research grant program, contact Alex Sedlacek at 202–219–1736, or e-mail alex_sedlacek@ed.gov.

The NAEP Guide, A Description of the Content and Methods of the 1994 and 1996 Assessments, revised 1996 edition, Nada Ballator, NCES 97–586. This gives a good overview of NAEP, and is available on-line at: http://nces01.ed.gov/pubsearch/infopage.idc?cid=97586XXXXX.

The NAEP Primer, 1995, Albert E. Beaton and Eugenio Gonzalez. The primer comes with a floppy disk containing a simplified NAEP database that modifies the sampling design to a random sample, which is easier to analyze. "We assume that the reader has a working knowledge of intermediate statistics including regression analysis and the analysis of variance. We also assume that the reader has a working knowledge of SPSS, a commonly available statistical system for mainframe and personal computers. The strategy is to get the user started quickly on a simplified database and introduce him or her to a few of the special features of NAEP." The Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, Chestnut Hill, MA 02167, 617–552–4521.

Using HLM and NAEP Data to Explore School Correlates of 1990 Mathematics and Geometry Achievement: Methodology and Results, Carolyn Arnold, NCES 95–697, available from the National Center for Education Statistics. This paper applies hierarchical linear models to the 1990 NAEP mathematics data to identify school, teacher, family and student correlates of overall mathematics achievement, and achievement on the NAEP subscale representing higher-level mathematics applications. In addition, this project developed new statistical software that facilitates the use of HLM with NAEP data.

Model-Based Methods for Analysis of Data from 1990 NAEP Trial State Assessment, Nicholas Longford, Educational Testing Service, NCES 95–713, available from the National Center for Education Statistics. This paper investigates the use of hierarchical linear models to estimate standard errors for student proficiency scores.

Note

Focus on NAEP is a series that briefly summarizes information about the ongoing development and implementation of the National Assessment of Educational Progress. The series is a product of the National Center for Education Statistics, Pascal Forgione, Commissioner, and Gary W. Phillips, Associate Commissioner for Educational Assessment. This issue was written by Alan Vanneman, of the Education Statistics Services Institute, in support of the National Center for Education Statistics. To order other NAEP publications, call Bob Clemons at 202–219–1690, or e-mail bob_clemons@ed.gov. The NAEP World Wide Web Home Page address is http://www.ed.gov/NCES/NAEP/.

United States
Department of Education
Washington, DC 20208–5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

FIRST CLASS